DOCUMENT RESUME

ED 356 469 CS 011 281

AUTHOR McEneaney, John E.
TITLE Neural Networks for Readability Analysis.
PUB DATE [93]
NOTE 25p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; Elementary Secondary Education; Higher Education; Measurement Techniques; *Networks; *Readability; *Readability Formulas; Reading Research; Regression (Statistics)
IDENTIFIERS *Neural Networks

ABSTRACT

This paper describes and reports on the performance of six related artificial neural networks that have been developed for the purpose of readability analysis. Two networks employ counts of linguistic variables that simulate a traditional regression-based approach to readability. The remaining networks determine readability from "visual snapshots" of text. Input text is transformed into a visual pattern representing activation levels for input level nodes and then "blurred" slightly in an effort to promote generalization. Each network included one hidden layer of nodes in addition to input and output layers. Of the four snapshot readability systems, two are trained to produce grade equivalent output and two depict readability as a distribution of activation values across several grade levels. Results of preliminary trials indicate that the correlation between visual input systems and judgments by experts is low, although, in at least one case, comparable to previous correlations reported between readability formulas and teacher judgment. A system using linguistic variables and numerical output correlated perfectly with a regression-based formula within the error tolerance established prior to training. The networks which produce output in the form of a readability distribution suggest a new way of reporting readability that may do greater justice to the concept of readability than traditional grade equivalent scores while, at the same time, addressing concerns that have been voiced about the illusory precision of readability formulas. (Three figures of data are included. Contains 45 references.) (Author)

Neural Networks for Readability Analysis

John E. McEneaney
Indiana University South Bend
South Bend, IN
Internet: JMcEnean@Indiana.edu
Voice: 219-237-4276
FAX: 219-237-4550

Running head: Neural networks for readability

Neural Networks for Readability Analysis

ABSTRACT

This paper describes and reports on the performance of six related artificial neural networks that have been developed for the purpose of readability analysis. Two networks employ counts of linguistic variables that simulate a traditional regression-based approach to readability. The remaining networks determine readability from "visual snapshots" of text. Input text is transformed into a visual pattern representing activation levels for input level nodes and then "blurred" slightly in an effort to promote generalization. Each network included one hidden layer of nodes in addition to input and an output layers. Of the 4 snapshot readability systems, two are trained to produce grade equivalent output and two depict readability as a distribution of activation values across several grade levels. Results of preliminary trials indicate that the correlation between visual input systems and judgements by experts is low although, in at least one case, comparable to previous correlations reported between readability formulas and teacher judgement. A system using linguistic variables and numerical output correlated perfectly with a regression-based formula within the error tolerance established prior to training. The networks which produce output in the form of a readability distribution suggest a new way of reporting readability that may do greater justice to the concept of readability than traditional grade equivalent scores while, at the same time, addressing concerns that have been voiced about the illusory precision of readability formulas.

Neural Networks for Readability Analysis

The concept of readability has had an important role in reading research, instruction, and the development of educational materials for many years. Traditionally, two different approaches have been taken in the measurement of readability: formula-based models and systems based on a standard set of reading passages or reading scale.

Formula-based measures of readability typically select a few text variables as a basis for estimating the difficulty of a passage. The text variables selected are usually based on a regression model developed to isolate the measures that most effectively account for variability. Fry (1977), for example, proposes a formula based primarily on sentence length and a syllable count. Raygor (1977) proposes a count based on sentence length and number of long words. Cohen (1975), Colby (1981), and Flesch (1950) on the other hand have propose counts based on the abstractness of the words used in the passage. In each of these formulas a regression equation links text difficulty (usually reported in grade equivalencies) to the text variables selected.

Readability formulas seem to work rather well as practical screening tools (Coupland, 1978; Rothkopf, 1980) and educators and publishers rely on them for a variety of purposes (Klare, 1984). They have, however, been criticized as promoting an illusion of precision (Fry, 1976; Klare, 1984) as a result of the regression approach. Typically, text variables go into a readability formula and a single readability output at a single grade level comes out (some formulas even report grade equivalencies in tenths of years). The problem is that the power of the mathematical framework can lead to an exaggerated sense of precision that goes far beyond, and may actually distort, the intuitions teachers have about what readability actually is and how it should be used in the classroom

(Resolutions of the delegates assembly, 1981; Fry, 1976).

The problems associated with reading formulas have led some reading researchers to propose an alternative approach to readability analysis based on reading scales (Carver, 1974, 1975-1976; Chall, Bissex, Conrad, & Harris-Sharples, 1983; Singer, 1975). A reading scale refers to a set of graded passages that serve as a standard against which individual passages are compared. In effect, reading scales are a formalization of "eyeball readability" (Singer even called his scale-based measure the Singer Eyeball Estimate of Readability - SEER). A teacher using a reading scale simply compares a passage to the set in the scale, looking for the passage in the scale that is most like the passage being evaluated. The readability of the passage in question is assumed to be the predetermined readability of the scale passage it is most like.

Studies of reading scales have shown, however, that the reliability of scale-based readability is user dependent, requiring subjects who are trained as "qualified" users. In a study that compared the Singer and Carver scales (Froese, 1980), one scale (Carver) was most reliable when multiple raters' results were averaged and the other (Singer) proved most reliable when applied by one specific rater who, apparently, was most adept in applying that scale. The increased flexibility and face validity of reading scales therefore appear to be purchased at the expense of reliability as a method since there is no guarantee different users will arrive at the same result from any given set of data.

It appears that both of the methods for determining readability that have been proposed suffer from shortcomings. The purpose of this paper is to describe a third approach, based on neural network technology, that is intended to integrate the benefits of both formula- and scale-based methods yet manage to avoid some of their limitations.

What is a neural network?

Neural networks are often referred to by cognitively-oriented researchers as Parallel Distributed Processing or PDP systems (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). PDP systems represent an approach to computing that attempts to make computers more brain-like in the way they operate and are constructed. PDP systems are based on the idea that information can be represented by the activation of a node or set of nodes in a network. But the real heart of PDP or connectionist systems is not the use of abstract nodes to represent information; the core of these models is that nodes are connected in a way that allows them to influence one another. Two kinds of connections are possible. An excitatory connection passes on activation to other nodes. An inhibitory connection serves to drain activation away. The introduction of an input to a PDP system consists of the external activation of input level nodes followed by the spread of activation and inhibition throughout the system. Problem solving is conceived of as the restabilization or "relaxation" (Rumelhart & McClelland, 1986b) of the system in response to the perturbation introduced by a stimulus.

In addition to an input level, PDP systems have an output level where each node or node aggregate represents a specific output value. It is common for connectionist models to also have one or more "hidden" levels of nodes between input and output layers. Hidden layers increase the complexity and corresponding power of the network.

Although PDP networks are direct descendants of Rosenblatt's (1959, 1962) perceptrons (simple networks with an input layer and a linear threshold output unit), they do not suffer from important limitations noted by Minsky and Papert (1969) in their now classic critique. For example, although simple two-layer associative networks like perceptrons are, in principle, incapable of solving the exclusive-or (XOR) problem (Rumelhart, Hinton, & Williams, 1986), the addition of hidden units in another (middle) layer changes the situation dramatically by providing an additional source of information (usually based on the overall output of <u>sets</u> of input units that is not available in the absence of a hidden layer.

Network A in Figure 1 illustrates a perceptron with two input units and one output unit with weight connections w1 and w2. Input units specify the information being provided to the system. A value of 1 means an input unit is activated; a value of 0 means the unit is not activated. Activation spreads from the input units according to weights that reflect the strength of connections between units which may be either excitatory or inhibitory. All of the activation flooding into the output unit is added up. If the sum of the activation coming into a unit exceeds that node's activation threshold (θ), the node turns on (and in multi-layer networks passes on the activation to other units).

Network A could, therefore, serve as a logical "and" gate by setting w1 and w2 to +1 and the activation threshold (θ) to 1.5. If the input provided to the system was (0,0), (0,1), or (1,0) the output unit would not come on since the activation coming in would never exceed the threshold θ. If the input was (1,1), however, the output would come on since $1+1 > \theta$. Solving the XOR problem however requires that the output unit come on when either one of the two inputs is activated ($w1 > \theta$ or $w2 > \theta$), but <u>not</u> come on if both input units are activated ($w1 + w2 < \theta$)! Given a perceptron's linear output unit, a solution is simply not possible.

Insert Figure 1 about here.

Networks B and C solve the XOR problem by adding one or more hidden units with negative (i.e. inhibitory) connection weights. In Network B, the (1,1) input will turn on the hidden unit but since the hidden unit has a connection strength of -2 with the output unit, the total activation at the output is 0 (1 + 1 - 2 = 0). Network C introduces 2 hidden units with negative connection weights between the input and hidden level. Inspection of the connections in network C reveals that although inputs of (1,0) and (0,1) will turn on the output unit, inputs of (0,0) and (1,1) will not.

Moreover, even if an additional layer were added to network A, the original perceptron learning procedure defined by Rosenblatt (1959, 1962) does not provide any way of adjusting the connection strengths to and from the hidden layer (i.e. the "credit assignment problem" identified by Samuel, 1959; 1967), a problem that played a central role in Minsky and Papert's (1969) critique and the subsequent decline of interest in perceptrons.

With the development of the back propagation learning rule (Werbos, 1974; Parker, 1982; Rumelhart, Hinton, & Williams, 1986), however, a mechanism became available for adjusting connection weights across one or more hidden layers by assuming all connections are at least partially responsible for errors. The degree to which any given output node deviates from the correct training pattern determines the extent to which connections are altered. The local error at any given output node, in turn, forms a basis for error computation at the next prior level. Through this step-wise "back propagation" of error all connections throughout the system are eventually adjusted. What this means for networks B and C in Figure 1 for example is that, although such networks might start out with randomly selected weights, there is a method of adjusting those weights using feedback about the correctness of responses that allows weights to be adjusted with whatever degree of precision desired; a method had been found that would allow these networks to learn the "right" weights.

Typically, a back propagation network includes an input layer, an output layer and at least one hidden layer. Layers are usually fully connected. Since its development, the back propagation rule has been a widely applied learning algorithm in PDP networks designed to address a variety of cognitively-oriented problems including speech perception (McClelland & Elman, 1986; Waibel & Hampshire, 1989; Waibel, Hanazawa, Hinton, Shikano, & Lang, 1989), word reading (Sejnowski & Rosenberg, 1987; Lacoutre, 1989; Seidenberg & McClelland, 1989), pattern classification (Gorman & Sejnowski, 1988), and vision (Lehky & Sejnowski, 1988).

Unlike traditional programs that are based on a set of instructions for executing a task, back propagation networks are <u>trained</u> to solve problems by repeated exposure to input/output pairs. The input corresponds to levels of activation for each of the input nodes in the network. The output pattern provides the "correct" activation level for each output node in the network. The difference between the correct output activation and the actual activation level for the output node is the basis of the error term that drives learning via back propagation. Input/output pairs are usually randomly ordered to avoid sequential learning effects. An input/output pair is sometimes referred to as a "fact" and each set of exposures to the entire data set is referred to as a cycle. One cycle through the data set, therefore means the system has been exposed to each input/output pair one time.

Initially, connections are randomly set. Early in the training process, therefore, the error terms across the output neurons are usually large and numerous connections are altered in fairly large steps. As the responses of the network improve, connections are altered in smaller and smaller steps. Over time, such systems tend to settle into a connection matrix that selects outputs with a minimum of error. When the system identifies all or most of the input-output pairs in the training data set within a specified tolerance, learning is complete and the system can then be tested with another independent data set. If the performance on the testing data set is acceptable, the system is ready for use. If the system's performance is not acceptable, the network must be trained further or redesigned.

The power and history of application of the back propagation algorithm (particularly in a variety of pattern matching tasks) seemed to suggest it might be an appropriate starting place to explore applications in readability assessment. In addition, numerous widely available development tools (McClelland & Rumelhart, 1988; California Scientific Software, 1990; NeuralWare, 1991) support back propagation networks so that any networks developed using this algorithm could be easily replicated by other researchers. Thus, although other, more specialized network architectures and learning algorithms could have been adopted, back propagation seemed a good choice and was used in all of the readability networks described below.

Neural network approaches to readability

The remaining portion of this paper describes six related back propagation networks that have been developed for the purpose of readability analysis. One network (FRY NET) simply implements a pre-existing readability formula (Fry, 1977). When this network is provided sentence and syllable counts, it generates a readability grade equivalent. Another network (FRY-ACTIVATION NET) takes input characteristic of the Fry formula (sentence length and syllables/word) but produces a readability distribution that can be interpreted as a probability statement about the readability of the text. The remaining (4) networks determine readability from fifty-word "visual snapshots" of the text. By using this approach, the text is treated as a visual pattern and readability assessment is a matter of pattern recognition. Although a pattern recognition approach to readability might seem unusual, it can reasonably be argued that this kind of approach is central to the use of reading scales and that pattern recognition as a basis of readability therefore has a precedent. Of these 4 snapshot readability systems, two are trained to produce numerical output (NUMBER NET 1 & 2) and two display readability distributions (ACTIVATION NET 1 & 2).

FRY NET: Fry formula input - grade equivalent output

The first of the networks described in this paper is FRY NET, a system that implements the Fry (1977) readability formula. Input to the FRY NET during training consists of numbers of sentences and syllables in a one-hundred word passage. Output from the system consists of a grade level readability score.

Training of the FRY NET consisted of repeated exposures to input-output pairs. On exposure to the two-valued input (numbers of sentences and syllables per one hundred words), the system generates a response which is then compared to the readability level that results from plotting the data on the Fry graph. If the system's guess falls within a specified range (10% or one about grade level) of the Fry readability of the passage the existing connections in the system are reinforced. If, however, the guess of the system falls outside of the acceptable 10% range, a signal is sent back through the network that alters the connections that have resulted in the output. This process of backwards adjustment of connections is the back propagation procedure described above.

The FRY NET was trained with a data set of 235 facts that were generated using the Fry readability graph. Input pairs were selected from the graph and the associated output was simply the Fry readability represented numerically at mid-grade level (third grade = 3.5, seventh grade = 7.5, etc.) Training required 31 cycles through the randomly ordered training data set for a total of 7314 exposures.

An independent testing data set consisting of 92 facts was developed using the Fry graph. Performance on the test set was 100% correct recognition within the 10% tolerance for error that was used in training. It appears that even with a fairly limited training data set, a network can be developed to effectively simulate the regression relationship embodied in a readability chart like that used by Fry. At the least, it would appear that neural networks based on linguistic variables offer a viable alternative to the standard regression-based approach.

NUMBER NETS 1 & 2: Picture input - grade equivalent output

It seems reasonable to ask, however, what if anything neural networks can offer beyond that provided by regression models. NUMBER NETS 1 & 2 provide a glimpse of one possibility; neural networks may provide a way to use the text itself as input (rather than abstract linguistic variables) by treating readability analysis as a form of pattern recognition.

Like the FRY NET, the number nets result in grade equivalent outputs. Unlike the FRY NET, however, the number nets take as their input not linguistic variables but relatively simple visual transformations of the text itself. Three-hundred-character samples of text were initially transformed into patterns of activation where each character or within-sentence punctuation assumes a value of +1, each space assumes a value of -1, and each end-of-sentence punctuation assumes a value of +10. The values assigned to letters, spaces, and punctuation are arbitrary in the sense that they have no meaning apart from the role they play within the network, although the value assigned to end-of-sentence punctuation (+10) does reflect an intentionally heavier weighting of end-of-sentence punctuation compared to letter input. It is possible that some other assignment of values could lead to different network performance. Informal experimentation with other value assignments, for example, suggests that assigning letters a value of 0, and spaces and punctuation positive values may help networks learn finer distinctions by lowering the overall activation within the network during processing.

Following assignment of activation values, visual displays were subjected to a gaussian smearing of their values in an effort to promote generalization by the network, a commonly used technique in training pattern recognition networks. The ultimate input to each number net is a 300 character transform (6 rows of 50 characters each) of text like that depicted at the bottom of Figure 2 which illustrates each stage of the data transformation process.

Insert Figure 2 about here.

NUMBER NET 1 was trained using snapshot inputs developed from passages randomly sampled from sets of graded reading materials from a number of commercially developed informal reading inventories (IRIs) including The analytical reading inventory (Woods & Moe, 1989), The classroom reading inventory (Silvaroli, 1990), The qualitative reading inventory, (Leslie & Caldwell, 1990), and The basic reading inventory, (Johns, 1991). Passages were word processed into electronic files and a special text processing program written especially for this research eliminated paragraph breaks, tabs and doubled spaces and reformatted each sample so that it would be readable to the networks. Patterns used to provide feedback to the networks were based on the grade levels assigned to the passages by the authors of the IRIs. NUMBER NET 2 was trained similarly to NUMBER NET 1 but employed passage samples taken from the Diagnostic Reading Scales (Spache, 1972).

Each number net was trained to a tolerance of 10% and then tested with a single testing data set that was developed independently of those used in training. Output from the number nets were grade equivalents similar to those produced by the FRY NET. NUMBER NET 1 correctly identified 34% of the testing facts within the 10% tolerance for error. NUMBER NET 2 correctly identified 36% of the testing facts within the established tolerance.

In addition, since the output of these networks is numerical, it was possible to evaluate their performance by calculating the correlation between their predicted output and the level determined by the authors. Output from both of the number nets were analyzed in this way resulting in an $R = 0.45514$ ($p < .05$) for NUMBER NET 1 and an $R = 0.10393$ ($p > .05$) for NUMBER NET 2. These correlations are well below what would be considered desirable although the output from NUMBER NET 1 is comparable to correlations that have been reported for some existing formulas (the Fog and Mugford formulas) when compared to teacher judgements of readability (Harrison, 1979). One possible explanation for the poor performance of NUMBER NET 2 is that the test data employed with this network was based on the IRI passages rather than samples from other passages in the Spache, but this "explanation" obviously undercuts the generality of the readability that is being assessed.

ACTIVATION NETS 1 & 2: Picture input - activation output

Although the number nets differed from the FRY NET by introducing a new visual way of representing information about text for the purpose of readability analysis, both the FRY NET and the two number nets produce numerical output that is interpreted as a grade equivalent, the traditional output of readability measures. There is, however, no reason that readability measures must produce numerical output and, as noted above, there are some drawbacks to the traditional grade equivalent output of readability measures. ACTIVATION NETS 1 and 2 explore two alternative ways of representing readability output in a non-numerical form.

The output layers in the activation networks consist of 10 nodes that represent readability from primer to the ninth grade level. Activation of each of these ten nodes is represented in histogram form with vertical activation bars like those depicted in Figure 3. When a grade level node is highly activated, a tall vertical bar appears over the corresponding grade level. When a node is only mildly activated, the vertical bar is short. Activation levels are interpreted as representing the extent to which a given text matches the grades represented across the output level. Another way these activation levels might be interpreted is as probability statements (Monteith, 1976) that the text is at a given grade level.

Since an activation distribution is the intended output of the activation networks, the training patterns provided to the system also take the form of activation distributions. Training patterns employed with ACTIVATION NET 1 are single fully activated output nodes with all other output nodes at zero activation. Training patterns used with ACTIVATION NET 2, on the other hand, are multi-grade "normally distributed" ranges of activations with means centered on the IRI grade levels of the passages as identified by the IRI authors. Activation distributions were created for each input by setting the node representing the mean readability level at full activation and 2 nodes to each side at progressively lower levels of activation. Distributions that extended beyond either end of the range (i.e. beyond primer and ninth grade) were truncated.

Networks that relied on visual input and a single fully-activated node as output proved to be very difficult to train. ACTIVATION NET 1, for example, required 38 hours of training with the input/output data set on a PC-based neural net simulator in order to achieve a 30% hit rate at a tolerance of 10%. Networks that produced activation distributions also required long training periods.

In order to account for the increased demand for accuracy imposed by output readability distributions, tolerance levels were reset to 40% for the distribution network. According to this tolerance level, each of the five grade level nodes must be activated to within 40% of the corresponding level in the correct target pattern. Although this enlarges the absolute level of activation discrepancy that will still satisfy the criterion, the number of possible outcomes is enormously increased since only one specific distribution out of all the possible distributions will be correct. This increased demand is reflected in the time required for the system to train. The use of the activation distribution as output also introduces complexities into the comparison between the network and other readability measures since it is not immediately apparent how comparisons should be made since the outputs differ in such a dramatic way.

Despite the relaxation of the tolerance condition, however, ACTIVATION NET 2 only achieved a 35% hit rate over a testing data set of 95 facts. It appears that the task demands exceeded the capacity of the network. Although a larger network on a PC (or on a more powerful computing platform) might lead to better performance, it is hazardous to generalize from the performance of one network to others. Although a successful network can be said to serve as an "existence proof" that a given task can be accomplished by a neural net, it is usually very difficult to articulate why complex nets do (or do not) work since they are founded upon the complex interactions of so many simple processing elements (McClelland & Rumelhart, 1981, p. 382).

A FRY-ACTIVATION HYBRID NET

The contrast between the excellent performance of the FRY NET and the relatively poor performance of the networks employing visual input suggested that perhaps it would be useful to consider a network that empl ,s more traditional linguistic variables such as syllables and sentences per one hundred words like the Fry yet generates output in the form of an activation distribution. Such an approach to network design would avoid problems visual input networks apparently had sorting through the enormous volume of information provided in the visual inputs yet still provide output in a more intuitive visual format that discourages the illusion of precision. A network was, therefore, developed that employed Fry input variables but was trained to provide output as activation distributions.

Since the FRY-ACTIVATION NET was being trained to produce activation distribution output, its training period was rather lengthy (approximately 10 hours). It did, however, successfully train to a much lower tolerance (.11) than was typical of other networks producing distribution outputs. This network was tested on both an independent data set of points randomly selected from the Fry graph and by another set of points specially selected to straddle grade equivalents. Performance on the random data set was 96% correct at the level that had been set for activation distribution networks (40%). The purpose of selecting points straddling grade equivalents on the chart was to examine whether the output distributions would reflect the ambiguous nature of the points' positions on the chart. Of the 24 ambiguous points selected 21 output distributions reflected that ambiguity in two adjacent nodes' activation levels exceeding .5 at the appropriate levels. Not only does the FRY-ACTIVATION NET provide accurate Fry readability, it also represents ambiguous points on the Fry chart in a straightforward visual manner (see Figure 3).

Insert Figure 3 about here.

Limitations and conclusions

The networks described in this paper range widely from fairly successful simulations of an existing readability formula to experimental designs whose input and output vary dramatically from past efforts and whose performances fall short of accepted standards. One response to these findings is simply to throw out systems that cannot meet acceptable performance standards but, it may be wise to reserve final judgement until the limits and capacities of networks like these are more fully explored. Neural networks have only recently begun to be applied to solving real-world problems. Their capacities as problem solvers are still largely unknown. It may be that readability is not appropriately treated as a pattern recognition problem but the data reported here are far too preliminary to make such a judgement at this time. Having raised a concern about too quick a judgment against neural network readability systems, however, it is also important to note possible limitations imposed by the choice of the PC platform.

It may be that PC-based simulation of neural networks simply does not provide the power needed to implement readability applications. Larger machines may be required to do justice to the concept of readability. All of the visual-input networks described in this paper translated text into fairly simple patterns of activation that may have obscured potentially important text features. Although, in principle, a neural network could be developed that responded to raw, unedited text, none of the systems described here do this, nor could they do this, given the limitations imposed by the hardware and software employed.

Given a commitment to desk-top systems (few teachers have access to a Cray on which to assess readability), it would appear that systems like those described here should focus on new ways of presenting output rather input. PC-based networks like the FRY-ACTIVATION NET that employ a limited number of linguistic variables but provide activation distribution output may be the most valuable short-term contribution systems like these can make to readability assessment. Hybrid systems that borrow both from traditional methods and the unique capabilities of networks offer educators new ways of reporting readability that seem to do greater justice to the concept and may help avoid its misinterpretation and misapplication in the classroom.

References

California Scientific Software. (1990). Brainmaker Professional. Grass Valley,CA: Author.

Carver, R.P. (1974). Improving reading comprehension: Measuring readability. (Final report, Contract No. N00014-72-C0240, Office of Naval Research). Silver Spring, MD: American Institutes for Research.

Carver, R.P. (1975-1976). Measuring prose difficulty using the Rauding scale. Reading Research Quarterly, 13, 116-32.

Chall, J., Bissex, Conrad, & Harris-Sharples (1983). Readability assessment scales for literature, science, and social studies. New York: McGraw-Hill.

Cohen, G.I. (1975). A new key to better writing. Paper presented at the 22nd International Technical Communication Conference. Anahiem, CA.

Colby, J. (1981). The i/e ratio - a measure of comprehensibility. Technical Communication, 28(1), 4-9.

Coupland, N. (1978). Is readability real? Communication of scientific and technical information, April 1978, 15-17.

Flesch, R.F. (1950). Measuring the level of abstraction. Journal of Applied Psychology, 34, 384-90.

Froese, V. (1980). Rauding and SEERing: Judging global readability. Perspectives on reading research and instruction. (29th Yearbook of the National Reading Conference.) Washington, D.C.: National Reading Conference.

Fry, E.B. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. Journal of Reading, 21, 242-52.

Fry, E.B. (1976). Crossfire: No precision formulas. The Reading Teacher, 29(7), 686-7.

Gorman, P., & Sejnowski, T. (1988). Learned classification of sonar targets using a massively parallel network. IEEE Transactions on Acoustics, Speech, and Signal Processing, 36, 1135-1140.

Harrison, C. (1979). Assessing the readability of school texts. In Eric Lunzer and Keith Gardner, The effective use of reading. London: Heinemann Educational Books.

Johns, J. (1991). Basic reading inventory (5th Ed.). Dubuque, IA: Kendall Hunt.

Klare, G. (1984). Readability. In Pearson, Barr, Kamil, & Mosenthal, Handbook of reading research. New York, NY: Longman.

Lacoutre, Y. (1989). From mean square error to reaction time: A connectionist model of word recognition. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), Proceedings of the 1988 connectionist models summer school. San Mateo, CA: Morgan Kaufman.

Lehky, S. R. & Sejnowski, T. J. (1988). Network model of shape-from-shading: Neural function arises from both receptive and projective fields. Nature, 333, 452-454.

Leslie, L. & Caldwell, J. (1990). Qualitative reading inventory. Glenview, IL: Scott Foresman.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

McClelland, J. & Rumelhart, D. (1986). Parallel Distributed Processing, Volume 2: Psychological and biological models. Cambridge, MA: MIT Press.

McClelland, J. & Rumelhart, D. (1988). Explorations in parallel distributed processing. Cambridge, MA: MIT Press.

Minsky, M., & Papert, S. (1969). Perceptrons. Cambridge, MA: MIT Press.

Monteith, M. (1976). Readability formulas. Journal of Reading, 19, 604-7.

NeuralWare. (1991). NeuralWorks Professional II/Plus. Pittsburgh, PA: Author.

Parker, D. (1982). Learning logic (Invention Report S81-64, File 1). Stanford, CA: Office of Technology Licensing, Stanford University.

Raygor, A.L. (1977). The Raygor readability estimate: A quick and easy way to determine difficulty. In P.D. Pearson (Ed.) Reading: Theory, research, and practice (26th Yearbook of the National Reading Conference). Clemson,SC: National Reading Conference.

Resolutions passed by delegates assembly. Reading Research Quarterly, 16, 16.

Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron. In Mechanisation of thought processes: Proceedings of a symposium held at the National Physical Laboratory, November, 1958. Volume 1 (pp. 421-456). London: HM Stationery Office.

Rosenblatt, F. (1962). Principles of neurodymnamics. New York: Spartan.

Rothkopf, E. (1980). Copying span as a measure of the information burden in written language. Journal of Verbal Learning and Verbal Behavior, 19, 562-572.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In David Rumelhart & James McClelland (Eds.) Parallel Distributed Processing, Volume 1 (pp. 318-362).

Rumelhart, D. & McClelland, J. (1986). Parallel Distributed Processing, Volume 1: Foundations. Cambridge, MA: MIT Press.

Rumelhart, D. & McClelland, J. (1986b). PDP models and general issues in cognitive science. In David Rumelhart & James McClelland (Eds.), Parallel Distributed Processing, Volume 1: Foundations. Cambridge, MA:

MIT Press.

Samuel, A.L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3), 210-233.

Samuel, A.L. (1967). Some studies in machine learning using the game of checkers, Part II. IBM Journal of Research and Development, 11(4), 601-618.

Seidenberg, M. & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96(4), 523-568.

Sejnowski, T. & Rosenberg, C. (1987). Parallel networks that learn to pronounce English text. Complex Systems, 1, 145-168.

Silvaroli, N. (1990). Classroom reading inventory (6th Ed.). Dubuque, IA: Wm C. Brown.

Singer, H. (1975). The SEER technique: a noncomputational procedure for quickly estimating readability level. Journal of Reading Behavior, 7, 255-67.

Spache, G. (1972). Diagnostic reading scales. Monterey, CA: CTB/Mcgraw-Hill.

Waibel, A. & Hampshire, J. (1989). Building blocks for speech. Byte, August, 235-242.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. (1989). Phoneme recognition using time-delay neural networks. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37, 328-339.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. dissertation, Harvard University, Cambridge.

Woods, M & Moe, A. (1989). Analytical reading inventory (4th Ed.). Columbus, OH: Merrill.

Figure Caption

Figure 1. A perceptron unable to solve the exclusive-or (XOR) problem (A) and two networks with one (B) and two (C) hidden units that are capable of solving the XOR problem. Connection weights (+w = excitatory connection, -w = inhibitory connection) are indicated beside lines representing connections; node thresholds are indicated within each node. (Adapted from McClelland & Rumelhart, 1988, pp. 124, 126, and 146.)

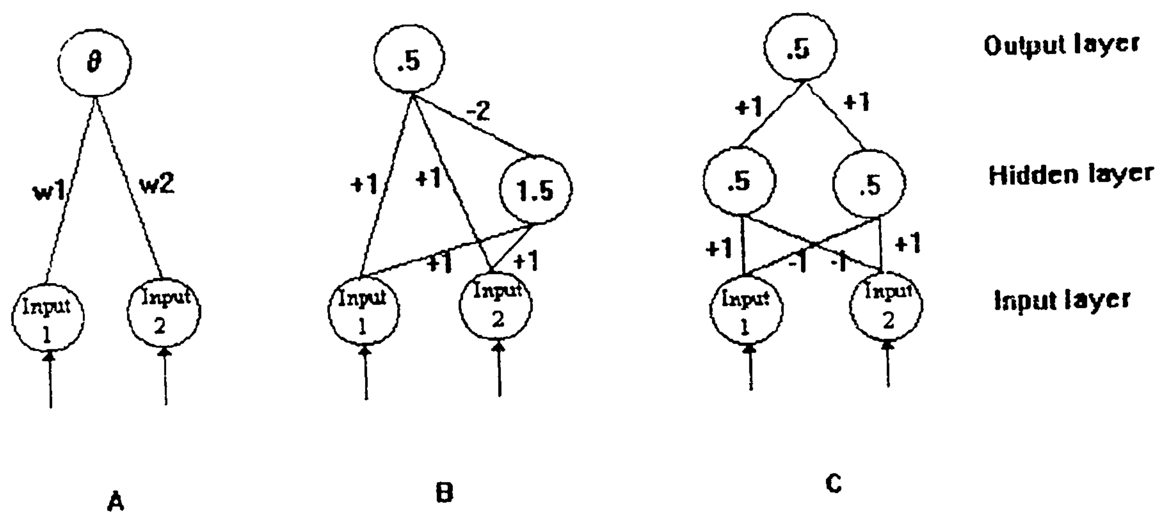
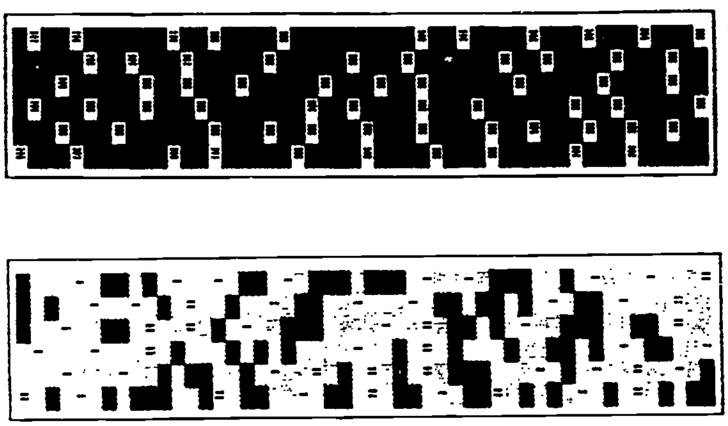


Figure Caption

Figure 2. Transform stages of data prepared for visual-input networks. A first grade passage and its distributed multi-grade output pattern is the top block. The activation transform of the passage is the middle block. The input following gaussian blurring is the bottom block. A chart for interpreting activation level symbols is at the right.

```
^A qrilb3
./He/wanted/to/find/something/to/eat./The/man/who/
lived/in/the/house/heard/the/mouse./He/knew/the/mo
use/lived/in/the/wall./But/he/didn't/mind./Then/on
e/day/the/man/decided/to/sell/the/house./He/loved/
the/old/house./But/it/was/too/big./He/put/an/ad/in
/the/paper./It/said,/"100/year/old/house/for/sale.
.6 primer one .6 two .3 three .1 four
```



	.900 ≤ value
	.633 ≤ value < .900
	.367 ≤ value < .633
	.100 ≤ value < .367
*	0.000 ≤ value < .100
.	0.000 = value
-	-.367 ≤ value < 0
=	-.633 ≤ value < -.367
≡	value < -.633

Figure Caption

Figure 3. Fry graph data points and their corresponding FRY-ACTIVATION NET outputs.